

DOCUMENT REVIEWED:	“The MPCP Longitudinal Educational Growth Study Second Year Report”
AUTHORS:	John F. Witte, Patrick J. Wolf, Joshua M. Cowen, David J. Fleming, Juanita Lucas-McLean
PUBLISHER/THINK TANK:	School Choice Demonstration Project; Department of Education Reform, University of Arkansas
DOCUMENT RELEASE DATE:	March 2009
REVIEW DATE:	May 28, 2009 [revised, June 29, 2009]
REVIEWER:	Casey D. Cobb
E-MAIL ADDRESS:	casey.cobb@uconn.edu
PHONE NUMBER:	860.486.0253
SUGGESTED CITATION:	Cobb, C. D. (2009). <i>Review of “The MPCP Longitudinal Educational Growth Study Second Year Report.”</i> Boulder and Tempe: Education and the Public Interest Center & Education Policy Research Unit. Retrieved [date] from <a href="http://epicpolicy.org/thinktank/review-MPCP-longitudinal">http://epicpolicy.org/thinktank/review-MPCP-longitudinal</a>

### **Summary of Review**

The study under review is the second-year evaluation report of the Milwaukee Parental Choice Program (MPCP), a publicly funded voucher program that allows low-income students in Milwaukee to attend secular and religious private schools in that city. Its primary finding is that there were no overall statistically significant differences in achievement growth in reading or math between MPCP and Milwaukee Public School (MPS) students over a one-year period. The study design and methods of analysis are sound overall. The relatively short duration of the treatment, however, raises questions about the usefulness of the findings.

### **Correction**

*On June 24, 2009, the authors of the report under review brought to our attention an error in the original May 28, 2009 review of their work by Casey Cobb. Professor Cobb has revised his review accordingly. As revised, the review includes an important correction. In the initial review, Professor Cobb made an incorrect assumption about the sampling procedures used in the study. As a result, he drew the incorrect conclusion that the external validity of the findings was tempered by a non-random sample of MPCP students, and hence a non-representative sample of MPCP schools. As the study authors have pointed out, the MPCP sample was in fact drawn randomly.*

## Review

### **I. INTRODUCTION**

The study under review is the second-year report of a long-term evaluation of the Milwaukee Parental Choice Program (MPCP). The MPCP is a publicly funded voucher program that allows low-income students in Milwaukee to attend private schools in that city. The program has been in operation since 1990 and has grown from serving 341 children in its pilot year to a current enrollment of around 20,000 students. MPCP voucher students may enroll in secular or religious private schools. The School Choice Demonstration Project at the University of Arkansas was commissioned to conduct a five-year evaluation of the program.

Four years ago, the passage of Wisconsin Act 125 called for several changes to the original MPCP legislation. One key amendment was the requirement for every school receiving voucher students to administer a nationally normed standardized test in reading, math, and science to its voucher students in grades 4, 8, and 10. Further, the law requires MPCP schools to submit scores from these tests and any other standardized tests to the School Choice Demonstration Project.<sup>1</sup> The Act also specifies that the evaluation report on the scores of a “representative sample of Choice pupils” (p. 4) on the Wisconsin Knowledge and Concept Examination (WKCE) and the Wisconsin Reading Comprehension Test (WRCT). The present *Second Year Report* represents a deliverable in keeping with that requirement.<sup>2</sup>

### **II. FINDINGS AND CONCLUSIONS OF THE REPORT**

The primary finding from this report is that

there were no overall statistically significant differences in achievement growth in reading or math between MPCP and Milwaukee Public School (MPS) students over a one-year period. The report offers a series of analyses that elicit comparisons between samples of MPCP and MPS students. The authors refer to those as “inter-sector” comparisons, implying that private school and public school treatment effects are under examination.

Mean post-test score comparisons did not reveal statistically significant differences overall. Nor did regression-adjusted comparisons controlling for grade level, race/ethnicity, gender, and mobility. The authors did report statistically significant negative effects for MPCP females in reading and positive effects for MPCP males in reading, relative to students of the same gender in the MPS sample. After accounting for the possibility of alpha inflation—the increased statistical chance of finding a significant result when conducting multiple comparisons—only the positive effect on MPCP males remained statistically significant. (The gender analysis is discussed in the following section of this review.)

Related to this achievement analysis was the finding that considerably more MPS students switched schools between October 2006 and October 2007 (the period studied). Because the disruption caused by changing schools has been associated in prior research with decreased test scores, the report makes note of this difference and offers some analyses.

In particular, 35% of MPS students in the sample switched schools within their sector, compared with only 8% of MPCP students. The relatively large numbers of MPS switchers apparently did so for both struc-

tural reasons (e.g., their school ended at a particular grade level) and non-structural ones (e.g., they left their present school for reasons unrelated to grade level). In terms of *between*-sector switching, 17% of the MPCP students moved to an MPS school in that one-year period. Roughly half of these students did so in either eighth or ninth grades, suggesting the switch was related somehow to attendance at an MPS high school. In contrast, less than 3% of the MPS sample transferred into an MPCP school.

As briefly noted above, the report includes additional regression analyses that controlled for school switchers (“mobility”), which still turned up no significant effects of the MPCP program, either positive or negative. The authors conclude that “the approximately similar growth rates between the two sectors is *not* driven by the disproportionate rate of school switching among MPS students” (p. 14).

### **III. THE REPORT’S RATIONALE FOR ITS FINDINGS AND CONCLUSIONS**

The authors were careful not to overstate the findings of a single-year study on achievement effects. Conclusions were tempered with several cautions and contingencies, almost to the point of raising questions about the utility of presenting any findings.

Only two significant findings emerged—both of them as a result of considering the interaction of gender with sector treatment. As noted above, MPCP males exhibited greater gains in reading relative to MPS males, and MPCP females exhibited lower gains in reading relative to MPS females. After the authors controlled for possible alpha inflation brought on by multiple comparisons, only the positive effect of MPCP males remained.

These findings raise questions as to why one sector might advantage males while the other might advantage females. The gender-specific comparisons across sectors were done without a clear rationale for doing so. Moreover, the significant finding among MPCP males was presented with little speculation on why this might be the case. An explanation of why the authors inspected these differences in the first place would have been in order and should be presented if the authors decide to proceed with similar interaction analyses in their future reports.

Moreover, this report does not offer an explanation of the 6.4-point predicted difference in reading outcomes between MPCP and MPS males (see Table 6, p. 17).<sup>3</sup> When statistically significant differences or effects are identified, it is usually helpful to the reader to express them in terms of a standard scale. But no effect sizes or expressions of the magnitude of any of the differences were presented in this report. In the authors’ defense, there really was not much to report on in terms of differences or effects in this particular study.

### **IV. THE REPORT’S USE OF RESEARCH LITERATURE**

Because this is an evaluation report, the research literature cited is primarily related to issues of research design and statistical modeling. The authors also refer to a few relevant school choice studies that examine effects on student achievement, most of which rely on econometric designs.

### **V. REVIEW OF THE REPORT’S METHODS**

The report employs quantitative models to tease out differences among groups. It presents a set of statistical analyses designed to detect sector effects. Each successive analy-

sis offers a more complex design to consider achievement differences between the MPCP and MPS samples. The analyses appear to be robust and rigorous. The writing is clear, and the explanations are helpful.

Key to this type of study is the ability to effectively identify a matched comparison group. The authors started with a random sample of MPCP students (more on the selection of this sample below), and made good-faith attempts to match this sample with a comparable sample of MPS students. A limited number of measurable, or observable, characteristics were used here to match MPCP students to a similar set of MPS students. The authors used students' neighborhood location, baseline test scores, and demographic indicators to create their comparison MPS sample. But of course this technique cannot ensure the two groups do not differ on other important characteristics that may influence their performance on the state tests or other outcomes of interest. These "unobservables" can bias the two groups and thus any inferences drawn upon their comparison.

As the authors acknowledged, a bias common to many school choice studies is *self-selection* bias. That is, the students and families that avail themselves of choice may differ systematically from students who do not partake in choice programs. The two groups could differ on unmeasurable variables such as academic motivation or family support of education—two factors that likely influence initial test performance as well as changes in test performance over time.

The authors claim that their use of a matching algorithm via propensity score techniques helps account for unobservable characteristics. In particular, their use of matching on neighborhood location, they assume, takes into account some degree of "home environ-

ment, parental education, and educational experiences for students and their parents from the same neighborhoods" (p. 6).

To be fair, this is probably about the best the report could do to effectively create a matched comparison group, and credit goes to the authors for using neighborhood location as a possible control. Where the report comes up short is in acknowledging and explaining the limited ability of this observable to control for the various unobservables that lead one group of students to pursue choice and others to not. Home environment is certainly related to academic performance; this is well established in the literature dating back to the Coleman Report. But to assume home location necessarily homogenizes families in terms of their likelihood to pursue or not pursue schools of choice is questionable. In fact, the logic flies in the face of the fact that students from the same home neighborhoods do indeed make different choices; some choose MPCP and others choose to remain in MPS.

Another important consideration, when comparing two groups over time, is the potential for unequal attrition from the two groups. Even assuming at the outset only random differences between groups, any subsequent attrition would have to continue in random ways, if the researcher is to maintain the integrity of the two-group comparison. If, for example, members who left the MPS sample systematically differed from members who left the MPCP sample, then the remaining samples are subject to bias. Ironically, the lead author of the report under review raised this concern more than 10 years ago in one of the early evaluations of the Milwaukee voucher program.<sup>4</sup>

Researchers are generally concerned if the absolute attrition numbers for any treatment group are too high; having too many subjects

exit a study raises questions about the external validity of its findings. (Of the remaining subjects in the two groups, what populations do they now represent?) There is also concern about the possibility that the two groups of subjects who left differ on a factor or factors related to the dependent variable. If, for instance, MPCP leavers are lower in academic ability or motivation or attendance relative to the MPS leavers, then the remaining participants in those two groups are now different in non-random ways that are related to the outcome of interest. If the MPCP sample outscores the MPS sample, this difference could be unduly attributed to sector, when such inferences potentially suffer from biased samples.

To their credit, the authors provide estimates of sample attrition—that is, the students they could not locate at the end of year two. Attrition was relatively moderate with 11% of MPS students and 7% of MPCP students disappearing from their respective groups. Importantly, there was little difference between missing and non-missing groups on several observable measures. Virtually no differences were evident in measured student characteristics between missing MPCP and missing MPS students.

Further, the authors elected to refresh the 2006-07 sample to recapture students who left the study. This is necessary for sustaining statistical power sufficient to detect effects across the five-year time period. This is certainly justifiable and common practice in the case of missing data in longitudinal designs. The departed group, however, should not be systematically different from the refreshed group. It is possible that those who left did so for reasons attributable to the treatment itself (i.e., the MPCP school).

Lastly, in lieu of experimental designs, regression techniques are often employed to discern treatment effects. This comparison-

with-controls strategy is probably the next best quantitative design to randomized clinical trials. But we should never assume multivariate correlational models of this sort are not without significant limitations. Regression does allow the researcher to examine the magnitude of the effect of group status (in this case, sector status), while simultaneously controlling for other factors known to be related to the dependent variable. The authors here attempt to statistically control for prior test score performance, race/ethnicity, the percentage of school switchers in each group, and gender. But the approach is always limited by the reality that unknown or unmeasurable covariates are not included.

## **VI. REVIEW OF THE VALIDITY OF THE FINDINGS AND CONCLUSIONS**

The internal validity of this study can be characterized by the degree to which one is confident that differences in achievement gains between sectors are attributable to sector type. In other words, is sector type related to growth in achievement? The answer appears to be “no.” There were no significant differences in achievement by sector group.

The authors appropriately and carefully issue cautions about drawing definitive conclusions from these findings. The most important admonition is that this study examines only one year’s worth of achievement. It is far more reasonable to expect that sector effects, if they existed, would surface over much longer periods of time—three to five years in some cases. There are also other student outcomes, such as attainment indicators (e.g., graduation rates), that are important to consider before drawing conclusions about the effects of sector type on low-income children. These data are being collected as part of the longer-term evaluation study.

It is also important to consider the extent to



which the report's findings generalize beyond the samples in the study (external validity). Before doing so, one must first determine what populations and settings the two samples represent. As is often the case, samples were drawn here *in lieu* of using the entire populations, due to the costs involved with collecting data from such large numbers of participants. The generalizability of study results relies on the samples being representative of the populations and settings that they characterize. The two samples in question here are the MPCP and MPS samples.

The MPCP sample for the *LEGS Achievement* study was constructed by drawing randomly from a list of September 1, 2006 participants who applied to and were accepted in the MPCP program. The sample was then stratified by MPCP enrollment in grades 3-8. A random sample of students was selected within each stratum, resulting in a total of 2,184 students. For the *LEGS Attainment* study, all ninth graders were added to the sample, bringing the combined total to 3,095 students.

From this list of 3,095 students for the combined *LEGS Achievement* and *Attainment* studies, 227 students presented as duplicates or were not on a Wisconsin Department of Instruction September 15, 2006 audited list of MPCP students. These students were dropped from the *LEGS* studies. The remaining 2,868 students were sent letters notifying them of the research. A relatively small number, 134 students (or 4.67%), declined participation. An additional seven students were no longer in grades 3-8 and were also dropped.<sup>5</sup>

For the MPS sample, randomized assignment of students was not possible due to the absence of lotteries or waiting lists. As noted above, the MPS sample was selected using a

multi-step algorithm that took account of individual students' neighborhood characteristics, prior achievement scores, and demographics. These matching procedures resulted in a sample that had similar observed characteristics to the MPCP sample. (Balance tests did indicate that the MPCP sample scored slightly lower on mean achievement measures in grades 3, 4, and 5; however, mean score differences in grades 6, 7, and 8 were negligible or absent). Overall, the MPS matched sample was more similar to the MPCP sample on test scores, gender, and race than a simple random sample of MPS students.<sup>6</sup>

Readers of this research should carefully consider the nature of the two samples from which inferences are drawn. First, consider the students in the two samples. The MPCP sample comprises students who opted to attend private schools, while the MPS sample consists of statistical matches on several key observables, save for one key variable: taking the private school choice option. One could argue that MPS sample *did* choose; in fact, they chose to attend their neighborhood public school. This is undoubtedly true of some, while others never actively considered an option, simply accepting the status quo option of attending MPS. The data are now quite dated, but there is little reason to question the continued relevance of Witte *et al.*'s (1994) *Fourth-Year Report* on the Milwaukee voucher program; that report found that, compared to MPS parents, the MPCP parents had better educations themselves and higher academic expectations for their children.<sup>7</sup> For these reasons, it may be most reasonable to say the current study speaks to the achievement effects of *low-income students who avail themselves of private school choice, in comparison with students who did not choose to seek a voucher but who were otherwise similar*. It should be obvious from that cautious statement that research of this

type, while very valuable, cannot offer definitive answers and should always be read as part of the larger body of research.

Research on school choice is a particularly politically charged arena. Accordingly, it is worth noting that the writing and analysis in this evaluation report are transparent to the reader and, more importantly, not particularly suggestive of bias either in favor or against school choice policy. The methodological approach appears sound and, notwithstanding the relatively minor concerns I raise in this review, adheres to rigorous principles of social science.

## **VII. USEFULNESS OF THE REPORT FOR GUIDANCE OF POLICY AND PRACTICE**

It is perhaps too early in the going for this *LEGS* evaluation report to have great utility to the field. Student achievement effects, should they exist, typically take more than one year to surface.

Moreover, the usefulness of the findings here depend in part on the nature of the

MPCP and MPS *schools* represented by these samples of students.<sup>8</sup> The study is designed to tease out school sector effects, or the effects of attending private or public schools. The fact that students self-select into the MPCP program introduces some level of bias that may conflate interpretations of sector effects. Although the authors took great pains to carefully identify a matched MPS sample, such non-experimental designs are more safely interpreted as “suggestive” of effects.

The longer, five-year evaluation will examine the longitudinal effects of MPCP participation on several attainment indicators, including promotion to the next grade, high school graduation, and college enrollment. The *Second Year Report* does not address these outcomes as it focuses on single-year achievement trajectories. Future reports that examine long-term impacts on academic achievement and attainment—and that clearly describe the private school population from which the MPCP sample was drawn—will likely be of more value to policymakers than this interim report.



## **Notes and References**

- <sup>1</sup> Source: [http://www.legis.state.wi.us/2005/data/lc\\_act/act125-sb618.pdf](http://www.legis.state.wi.us/2005/data/lc_act/act125-sb618.pdf)
- <sup>2</sup> Witte, J.F., Wolf, P.J., Cowen, J.M., Fleming, D.J. & Lucas-McLean, J. (2009). "The MPCP Longitudinal Educational Growth Study Second Year Report." School Choice Demonstration Project, University of Arkansas, Department of Education Reform. Retrieved May 27, 2009, from [http://www.uaedreform.org/SCDP/Milwaukee\\_Eval/Report\\_10.pdf](http://www.uaedreform.org/SCDP/Milwaukee_Eval/Report_10.pdf)
- <sup>3</sup> If the standard deviation of the WKCE is around 50, as some of the baseline data seem to suggest, a scale score difference of 6.4 amounts to .128 standard deviation units. Effect sizes of this magnitude are considered small. See: Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- <sup>4</sup> Witte, J.F. (1997, January). *Achievement effects of the Milwaukee Voucher Program*. Paper presented at the 1997 American Economics Association Annual Meeting, New Orleans, January 4-6, 1997.
- <sup>5</sup> See Appendix B of the report for an account of sampling procedures.
- <sup>6</sup> For a detailed description of sample construction, see Witte, J.F., Wolf, P.J., Cowen, J.M., Fleming, D.J. & Lucas-McLean, J. (2008, February). "MPCP Longitudinal Growth Study Baseline Report." School Choice Demonstration Project, University of Arkansas, Department of Education Reform. Retrieved June 28, 2009, from [http://www.uark.edu/ua/der/SCDP/Milwaukee\\_Eval/Report\\_5.pdf](http://www.uark.edu/ua/der/SCDP/Milwaukee_Eval/Report_5.pdf)
- <sup>7</sup> See Witte, J.F., Thorn, C.A., Pritchard, K.M., & Clairborn, M. (1994). *Fourth-Year Report: Milwaukee Parental Choice Program*. Madison, WI: The Robert La Follette Institute of Public Affairs, University of Wisconsin-Madison.
- <sup>8</sup> It would be helpful for policymakers and practitioners to know if sector effects are distributed equally across sectors, or if certain private or public schools appear to induce greater influences on achievement.

---

The Think Tank Review Project is made possible by funding from the Great Lakes Center for Education Research and Practice.